Tying Odysseus to the Mast: Using the Research on Online Reading to Resist the Call of New Educational Technologies

Randy Connolly
Mount Royal University (Canada)
rconnolly@mtroyal.ca

Abstract

For much of the past fifteen years, educators have been encouraged to adopt (and adapt to) online technologies. While educators have often had reservations about new technologies replacing old technologies, educators do not always have the opportunity to fully examine the evaluative research on new education technologies. Indeed, quite often evaluating a new education technology is only possible after it has been adopted. Thus teachers are frequently initially encouraged to adopt a new technology not by research demonstrating its merits, but by a series of metaphors or analogies comparing the new unknown technology to an old and trusted one. Yet even if research ultimately emerges that is critical of the new technology, by then it is too late; institutional inertia and interest politics make it difficult to un-adopt a technology.
A complicating factor in evaluating the research in this area is that the relevant research approaches are quite often not those typically used by education researchers. This paper attempts to address this last problem by providing an overview of the different research approaches that can be used to evaluate the effect of web-based technology on reading and learning, and thereby broaden the evaluative perspective of educators. This paper, which is based on a book in progress, shows that if one does branch out of the usual education research approaches, one will see that there is in fact a great deal of evidence that shows that the replacement of paper reading with the online consumption of text is a dangerous siren call and one that educators would be wise to resist.

In Book XXII of Homer's *Odyssey*, the hero Odysseus recounts his adventures after leaving the island of Circe, perhaps the most famous of which is the one that gives this paper its title, the island of the Sirens. Forewarned by Circe about the dangers of hearing the impossible-to-resist song of the Sirens, Odysseus has instructed his crew member to stuff their ears with wax. But being the archetype of the curious seeker after knowledge and experience, Odysseus does not do the same. Aware that he will be unable to resist temptation, he instructs his crew members to bind him tightly to the mast, so that he can hear the Sirens' song, but will be unable to act upon it. In this story, Homer supplied a timeless metaphor about the temptation of knowledge, one that is paternalistic for sure, but one which still has relevance today.

Since the popularization of the web in the mid-1990s, educators have often found themselves facing a similar dilemma as Odysseus. Like Odysseus, most educators are curious, knowledge- and experiencing-seeking creatures. But unlike him, we do not have the advantage of Circe's warning about upcoming dangers. At best we only have research to inform us about the value or problems with educational technology. Unfortunately, educators do not always have the opportunity to fully examine the evaluative research on new education technologies. Indeed, quite often evaluating a new education technology is only possible *after* it has been adopted. Thus educators are frequently encouraged to adopt a new technology not by research demonstrating its merits but by a series of metaphors or analogies comparing the new unknown technology to an old and trusted one (such as comparing the web to the printing press). Yet even if research ultimately emerges that is critical of the new technology, by then it is too late. Institutional inertia, financial investment, and interest politics make it difficult to un-adopt a technology.

A complicating factor in evaluating the research in this area is that the relevant research approaches are quite often not those typically used by education researchers. This paper attempts to address this last problem by providing an overview of the different research approaches that can be used to evaluate the effect of web-based technology on reading and learning, and thereby broaden the evaluative perspective of educators. This paper, which is based on a book in progress, shows that if one does branch out of the usual education research approaches, one will see that there is in fact a

troubling amount of evidence that shows that the replacement of paper reading with the online consumption of text is a dangerous siren call and one that educators would be wise to resist.

There is no centralized agency that records what people do online. To understand how people use the web, we must instead rely on a variety of indirect measures. Each of these has certain strengths and limitations. The first and least reliable measure is anecdotal evidence. It is rather astonishing just how often people are content to universalize from the one or two examples they may have observed. When we have no other evidence available, it is not surprising that we would do so. It is much less acceptable when academics or journalists do this even when better evidence is readily available. It is not uncommon to encounter a story about digital culture that begins with the author recounting what his or her friends or children are doing with the internet, and then extrapolating that out to a general society-wide trend (for instance, see [1]). If the original article was in a well-known publication or blog, other writers re-report these “findings” and then others re-re-report them, and pretty soon, a new “truth” about the internet has been created.

Anecdotal evidence can appear especially compelling if it reinforces an existing stereotype. We often see this in the many contemporary pronouncements that claim that the younger digital generation is adroit at mastering new technologies because, well, because they are young. Of course, we all have stories of older people laughably clueless in the presence of new technology while their children or grandchildren have no such problem. But these clichés about a digital generation “have been subjected to little critical scrutiny, are under theorized, and lack a sound empirical basis” [2]. When these clichés are subjected to the exposure of empirical testing, one generally finds that these anecdotal-based beliefs not only lack real evidence, but indeed are often even false [3].

For this reason, we should not rely on anecdotal stories or the half-truths that are spun out from them. Instead, when talking about the web, we should endeavour to rely on better, more empirically-grounded evidence. One of the most popular approaches to obtaining such evidence is the use of surveys. For instance, the influential PEW Research Center's Internet & American Life Project base their findings almost exclusively on surveys. These surveys can be online; however, since it is difficult to reliably control for variables such as age, race, income, and nationality with online polls, most reputable surveys are conducted in person, via mail-outs, or via the telephone. (The problem with online polls is more precisely called a sampling bias, which refers to the statistical unreliability of a non-random population sample).

While surveys provide interesting evidence, they are not conclusive on their own. For instance, in a survey-based study of faculty and graduate students, 80% indicated that they were reading less but scanning more, while 50% indicated that they had less sustained attention when reading and did in-depth reading much less frequently than they once did [4]. But how reliable is this evidence? What does it actually show? Humans are remarkably subjective creatures and we sometimes do not accurately reveal our actions or attitudes, especially if we feel somewhat embarrassed by them. Statisticians refer to this phenomenon in general as a response bias, or, more precisely, a social desirability bias. As a consequence of the social desirability bias, surveys can be excellent vehicles for capturing what we claim to feel about something but not necessarily what we actually feel. That is, surveys capture attitudes but are flawed for capturing actions.

One instructive example was a study [5] that timed how long professors took to read an academic article online and then asked them to estimate the length of time they had spent reading it. The study found a tremendous variance between the actual online reading times (generally about one to two minutes) and their self-reported reading times (generally between ten to fifteen minutes). Being professional readers, academics in this study no doubt felt that they should be spending about 10 minutes to read a paper. Yet clearly the self-reporting of academics was a significantly unreliable indicator of their actual behaviours due to a response bias.

A better way to find out how people use the web is to observe them when they use the web. Outside of, say, North Korea, this is generally impossible to do without the subject’s approval. What we can do is to round up some volunteers and bring them into a computer laboratory and then observe what they do. While this approach is certainly more accurate than surveys, one might argue that people most likely behave somewhat differently in an observed laboratory than when alone. We know from server records that a significant percentage of users regularly visit pornographic sites (some estimates put

that percentage at around a third of all users); it would take a pretty brazen individual in a laboratory to do so knowing that he/she is being observed. In open browsing sessions in an observed laboratory, participants will be much more likely to visit news sites than they would in the privacy of their home or office.

One way to avoid response biases in the study of internet usage in the laboratory is to give participants explicit goals and then evaluate in some way what they do, either by observation or by some type of testing after completion of the task. Most educational evaluation of technology falls into this style of research, which can be particular tricky when it comes to assessing the effect of a technology on reading abilities and/or learning. How important are the subjects' post-testing preferences? How much can we trust evaluations of comprehension (which usually translates to testing the subjects' recall of what they read)? Can we evaluate possible unique connections between knowledge that a new technology might be delivering?

There is certainly conflicting results within the very vast literature on the educational evaluation of internet-based technology. The fact that some research in this area has found that "the net total effect of the web is actually to reduce learning compared to print presentation" [6] and for "the more critical and creative skills … experience and exposure to [online] information seem to have a negative effect on the user's performance" [7], might make us indeed try to resist the siren call of online reading, but taken by itself, we should not make too many conclusions from this type of research, especially given the existence of countervailing findings.

Another type of laboratory-based research approach to examining how people use the internet is more commonly found in computing disciplines. This type of study is more generally referred to as "usability analysis." Many commercial web sites hire specialty firms to perform this type of analysis. The test subjects will be given a series of goals, such as "find the contact phone number for the company," "find the price of X," or "using the following customer information, try to buy these two products." The test subjects are observed, either directly or by video recordings or screen-capture recordings. By examining the recordings, the analyst can provide the web site owner with valuable information about time to completion, failure percentage, degree of lostness, and a variety of other measures. The goal of such analysis is to create a better experience for the user. Over a decade of rigorous, well-funded usability testing has provided an insightful picture of how people use the web – one that is significantly more reliable than surveys.

In recent years, laboratory usability testing has been revolutionized through the use of eye-tracking technology. This particular technology has been around since the 1970s and has played an important role in the transformation of newspaper design in the 1990s. Most modern eye tracking systems use video images of the eye, typically by recording the reflections of infrared light off the cornea. Contemporary eye-trackers now look much like over-sized plastic sunglasses that are missing their lenses. Very small cameras within the frames are then able to track the movement of the eyes. The data generated by eye tracking allows the researcher to understand what parts of the scene or page or screen received the eyes' attention.

Eye trackers generate a lot of data. Sophisticated software is required to process and aggregate this data. A variety of visual forms are used, such as heat maps and fixation bubbles, to represent what parts of the page or screen received the eye's gaze. The price drop in this equipment has allowed a wider variety of research in both web usability and reading research that has given us a much fuller picture of how we read online.

This technology provides an unparalleled picture of what actually happens when subjects consume text on web pages. *It provides very strong evidence that people rarely read online but engage in a highly-optimized, very-fast scan of web pages for answers to specific questions*. We now have direct and clear evidence that users' eyes spend very little time on a web page's text and only glance at a small fraction of the text [8]. One of the most experienced researchers in this area summarizes his eye-tracking research by noting that the only online text that users reliably read are the first eleven characters (about two words) of headings [9].

Eye tracking, one might argue, still suffers the same drawbacks as all laboratory studies, namely, that the subjects are in an "unreal" situation, and as a consequence the gathered data might not accurately

reflect "real" usage. Having a sufficiently large sample size can perhaps ameliorate this problem. Another way to address this problem is to convince test subjects to willingly allow their browser sessions to be remotely observed across longer time periods through the voluntary installation of a type of monitoring software that records everything they do in their browser, and then sends the data back to the research team. While this approach is not fully free from response biases (since the subjects are aware of it and indeed volunteered for it), it can provide the closest we can get, short of illegal surveillance, to knowing how individual users really use the web across time. This approach has given us findings that are consistent with those from eye tracking studies, namely, that almost all web page visits are exceptional fast and the text consumption on those pages is no more than cursory. One study using this approach [10], found that the average stay time for any given web page is ten seconds, a finding that has been replicated in other studies [11].

The only way to legally overcome the limitations of all laboratory-based research in this area is to make use of server records. Every time you visit a web page, there will often be dozens or even hundreds of requests for resources, such HTML content, CSS and Javascript files, as well as many image files. These requests are serviced by computers known as web servers. These servers can be configured to record or log information from these requests. When aggregated, these server logs can contain millions of requests. While individual behaviours are lost in this aggregation, the analysis of such large data sets certainly provides an alternate insight into user behaviours.

So what does this server data tell us? It in fact provides valuable corroborating evidence for the fast-scanning behaviours mentioned earlier. One study that examined 50 million server records found that the median stay time for Yahoo users in 2009 to be 12 seconds and fewer than 10% were for longer than 90 seconds [12]. Researchers examining requests for the ScienceDirect academic database found similar results [13].

So, to summarize, if one wishes to evaluate the educational impact of web-based technologies, one can make use of evidence from surveys, from laboratory observation, from eye-tracking studies, from monitoring software, and from server logs. Each has its own strengths and weaknesses. But when combined, these evidence-based examinations of web usage allow us to truly understand what people really do when they use the web.

Taken as a whole, evidence using all these different methodologies paints a remarkably consistent and clear picture. It shows that when we consume text, we are in fact engaging in a highly-optimized gathering of information that should not be called reading, but should more accurately be called scanning. If we want ourselves, our students, or our children to read, then do not ever expect web page viewing to be a replacement for reading because it most certainly is not reading. As Hillesand concluded from his 2010 study of how readers "handle" books, web pages, and electronic reading devices, "web browsers should be used for what they are good at – presenting overviews and accessing information through links and search functions" [14]. But for actual reading, then print (or perhaps eventually e-book devices) should be used.

Like Odysseus, we educators are curious and often eager to experiment with new learning approaches made possible by new technologies. And while we are unable to avail ourselves of a Circe to warn us of dangers ahead of time, we should be willing to make use of a wide range of different research approaches when evaluating the benefits of computing technologies to education. In the case of online reading technologies, we would be wise to listen to the research findings in disciplines outside of education, which provides strong evidence that while new reading technologies may be seductive, they are no replacement for books.

References

[1] Tapscott, D. (2009). *Grown Up Digital*. McGraw Hill.
[2] Bennett, S., *et al*. (2007). "The 'digital natives' debate: A critical review of the evidence." *British Journal of Educational Technology* 39 (5).
[3] Selwyn, N. (2009). "The digital native – myth and reality." *New Information Perspectives* 61 (4).
[4] Menzies, H. and Newson, J. (2006). "No Time to Think?" *Academic Matters* (Winter).
[5] Nicholas, D. *et al* (2008). "Viewing and reading behavior in a virtual environment." *ASLIB Proceedings: New Information Perspectives* 60 (3).

[6] Eveland, W. P., and Dunwoody, S. (2002). "An Investigation of Elaboration and Selective Scanning as Mediators of Learning from the Web Versus Print." *Journal of Broadcasting & Electronic Media* 46(1).
[7] Eshet-Alkalai, Y. and Chajut, E. (2009). “Changes Over Time in Digital Literacy.” *CyberPsychology & Behavior* 12 (6).
[8] Shrestha, S., and Owens, J. W. (2009). “Eye Movement Analysis of Text-Based Web Page Layouts,” Usability News (11).
[9] Nielsen, J. (2009). “First 2 Words: A Signal for the Scanning Eye.” *http://www.useit.com/alertbox/nanocontent.html.*
[10] Weinreich, H., et al. (2008). "Not quite the Average: An Empirical Study of Web Use." *ACM Transactions on the Web* 2(1).
[11] Kelly, D. and Belkin, N. (2001). “Reading Time, Scrolling and Interaction: Exploring Implicit Sources for User Preferences for Relevance Feedback”, *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*.
[12] Kumar, R., and A. Tomkins. (2010). "A Characterization of Online Browsing Behaviour." *Proceedings of the 19th International Conference on World Wide Web*.
[13] Huntington, P. *et al*. (2008). "Website Usage Metrics: A Re-Assessment of Session Data." *Information Processing & Management* 44(1).
[14] Hillesund, T. (2010) "Digital Reading Spaces: How Expert Readers Handle Books, the Web and Electronic Paper." *First Monday* 15(4).